Data Stack Evaluation Checklist

This checklist is designed to support real-world evaluation of data stacks in professional settings. It is intended to guide discussion and decision-making, not dictate tooling choices. Not every item needs to be a strict requirement, but unanswered questions often indicate hidden risk.

1. Purpose & Context

☐ What problems is this stack meant to solve right now?
☐ Who are the primary users (data engineers, analysts, ML, business)?
☐ What scale assumptions are we making (data volume, team size, growth)?
☐ What constraints exist (cost, compliance, latency, skills)?

2. Clarity & Usability

☐ Can a new engineer understand the system within 30–60 days?
☐ Are naming conventions and directory structures consistent?
☐ Is the data flow easy to follow end to end?
☐ Are core workflows documented?
☐ Is the happy path obvious?

3. Ownership & Responsibility

☐ Is ownership defined at each layer (ingestion, transformation, consumption)?
☐ Do teams know who responds when data is incorrect or stale?
☐ Are escalation paths clear during incidents?
☐ Are responsibilities documented or implicit?

4. Observability & Reliability

☐ Can we easily tell if data is fresh?
☐ Are partial failures visible?
☐ Are checks in place for schema changes or missing data?
☐ Can issues be debugged without deep system archaeology?
☐ Are failures surfaced early enough to act on?

5. Change & Evolution

☐ Can schemas evolve without breaking downstream consumers?
☐ Is raw data isolated from business logic?
☐ Can new sources be added incrementally?
☐ Are migrations manageable or all-or-nothing?
☐ How difficult is it to undo a bad decision?

6. Human Cost & Cognitive Load

☐ How much undocumented knowledge is required to operate the system?
☐ Can engineers make changes confidently?
☐ Is the system understandable during incidents?
☐ Does the stack reduce or increase cognitive load?

7. Long-Term Sustainability

☐ Is this stack operable by future team members?
☐ Are we dependent on a few key individuals?

☐ Can the system evolve with organizational change?
☐ Are operational costs visible and understood?

8. Tradeoff Awareness

☐ What are we explicitly optimizing for?
☐ What tradeoffs are we consciously accepting?
☐ What risks are acceptable?
☐ What risks are unacceptable?